Internet On-Ramp

Searching the Internet

Robert M. Horton - Department of Dermatology
University of Minnesota, Minneapolis, MN, USA
Internet: horto005@maroon.tc.umn.edu

"We are in great haste to construct a magnetic telegraph from Maine to Texas; but Maine and Texas, it may be, have nothing important to communicate." - Henry David Thoreau, *Walden*, ca. 1854.

There's always a skeptic. People can look at the amazing technical marvels that comprise the Internet and conclude that it may be pretty, but question its usefulness. And don't you have experiments to do, anyway?

Undeniably, there are many Internet sites of questionable utility. There are archives of David Letterman's Top Ten lists (<http://www.cbs.com/lateshow.tlilist.html>). You can get the latest "Dilbert" cartoon (<http://www.unitedmedia.com/comics/dilbert/>) or see the classic LIVE pictures of pet fish (<http://www1.netscape.com/fishcam/fishcam.html>). You can even find photos of people's spouses (<http://155.41.115.114/leach.html#wedding>) and of favorite vacation spots (<http://134.84.47.3/Pages/Vacations/glacier95.html>). And yes, you can find dirty pictures, jokes and discussion groups. (Hey, don't look at me!)

But there are also Net resources that can be tremendously useful in research, and can save you time and even grant money. There are general approaches to using the Net more effectively to find what you want, and to avoid getting lost in Cyberspace.

A first rule is to "surf" with a purpose. If your purpose is to familiarize yourself with Internet resources, and you wander out among the biology-related sites, that's one thing. But you must exercise discipline to avoid the temptation of unrelated topics (are you STILL reading "Dilbert"?).

One of the best ways to find useful sites is for someone to tell you about them, and we'll do our best in the **Internet On-Ramp** to bring such sites to your attention. People also identify relevant Web sites in BIOSCI/bionet newsgroups. Newsgroups also provide an opportunity to ask your colleagues to point you toward sites on a particular topic.

Other ways to find particular sites are "directories" and "search engines". With directories you ask, "what topics are available on these pages?"; with search engines you ask, "what pages are available on this topic?" Directories are collections of links to sites grouped by subject. For almost any field you can think of, someone has probably set up a Web directory page—your only problem is to find directories relevant to your needs. *Net Directory* on Netscape's toolbar, for instance, takes you to one called *Excite*.

A search engine is a program that lets you tell it what to look for, and it ferrets out sites it deems related to your topic. Several directories, including links to search engines, are listed in the Net Navigation Hint.

Browsing a directory may inspire you to use a search engine to locate other resources. Many directories include links to search engines, and some even have search functions built in. Also, many search engines track popular sites accessed through their service and maintain a list of these "hot" sites.

news.group.news

News.group.news goes to cyberspace and back to gather Net news for you. We emphasize practical, methods-related issues that are discussed in many newsgroups and also feature issues of general interest to biological scientists. But be warned! On this page of *BioTechniques*, we don't review it—we just report it!

You can't believe your eyes because your favorite protein is exhibiting **SDS-resistant dimerization** (or tri-, tetra- or hexamerization) following electrophoresis under denaturing conditions? This behavior is considered appropriate if the protein is an integral membrane protein, such as glycoporphin A or the bacterial porins, and has been attributed to properties of the transmembrane domain. Netters recommend the following reading on the subject: Yu et al. 1979. A major outer membrane protein (O-8) of *Escherichia coli* K-12 exists as a trimer in sodium dodecyl sulfate solution. *FEBS Lett.* 100:71-74 and Lemmon et al. 1992. Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry* 31:12719-12725.

Is your **dam methylation-sensitive restriction enzyme**, *BsaBI*, refusing to cut your non-*dam*-methylated plasmid? Methylase mavens on the Net suggest that if you are doing your *BsaBI* digestion at 60°C (as you should) then you ought to (i) confirm that your plasmid is indeed non-*dam*-methylated by digesting it with, e.g., *Mbol* and (ii) determine whether your *BsaBI* is active by using it to digest suitable control DNA.

Broad bands on Northern blots, "tilted" due to differential migration of molecules present in the upper and lower portions of the sample, were plaguing one **bionet.molbio.methods-reagents** newsgroup user. Advised by helpful contributors to completely equilibrate the gel in running buffer, he posted back to the group that this advice, coupled with a reduction in the concentration of formaldehyde—from 2.2 to 0.2 M—in the gel and buffer, completely eliminated the problem.

In response to a request for protocols and recipes for "homemade" reagents for use in **silver staining of DNA** in gels, several Netters posted their personal procedures and acknowledged the following publications as their starting points: Bassam et al. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196:80-83 and Gotlieb and Chavko. 1987. Silver staining of native and denatured eucaryotic DNA in agarose gels. *Anal. Biochem.* 165:33-37.

So - you love your little **zebrafish (*Danio rerio*) fry**, but culturing those *Paramecium* for them to eat is very time-consuming and you can hardly wait for the day when they're big enough to eat brine shrimp? According to postings on **bionet.organisms.zebrafish**, if you like to make your own baby food, it's very simple to culture all the *Tetrahymena* they can eat. But flake food and liquid drop food work fine, too.

BioTechniques Editorial Department

Internet On-Ramp

Net Navigation Hint: Internet Catalogs

Internet catalogs or "yellow pages" help you browse more effectively by categorizing Net resources. Instead of looking randomly through sites, you can start with a collection of sites related to biological science. The Uniform Resource Locators (URLs) below take you to biologically related sections. You can usually truncate a URL to go up to more general levels (i.e., the *Excite main level* and *search engine*, is at URL <http://www.excite.com>)

Excite <http://www.excite.com/Subject/Science/s-index.h.html>

Yahoo! <http://gnn.yahoo.com/gnn/Science/Biology/>

The World Wide Web Virtual Library - BioSciences <http://golgi.harvard.edu/biopages/>

"Gopher Jewels" catalog of gopher sites - Biological Sciences <http://galaxy.einet.net/GJ/biology.html>

"Pedro's BioMolecular Research Tools" directory of molecular biology sites, programs for sequence analysis, etc. http://www.public.iastate.edu/~pedro/research_tools.html

"Shareware.com" directory; searchable links to over 160,000 shareware and free programs <http://www.shareware.com/>

Finally, a catalog of many World Wide Web (WWW) search engines can be found at: <http://www.indiana.edu/~librcsd/search/>

You can also use your browser's "find" function to locate a particular term in a big document. And *Net Search* on Netscape's toolbar takes you to a search engine with links to other search engines.

Here are some specific examples of using a search engine:

Simple searches are straightforward. For example, if we are interested in the Ebola virus, we could probably just press Netscape's *Net Search* button, and type in Ebola. Since this word does not arise frequently in casual conversation, documents that use it probably are concerned with the virus.

For a more complicated example, let's look for free software that lets you analyze or process video images. This could be useful, letting you trade in your old Polaroid® DNA gel camera. But if you don't know the name of any such program, how do you find it?

This time let's use the *Alta Vista* search engine (<http://altavista.digital.com/>) from Digital Equipment Corporation (they use it to show off their fast Alpha chip). It is fast indeed, and it lets you enter complicated search criteria. Different search engines may use slightly different syntax, but they usually link to pages that tell you how to use them.

In *Alta Vista*, click on the box marked "help" to get instructions. We'll use the *Advanced query* button to enter our sophisticated selection criteria.

We could type in *image analysis*, but what if the page we want calls it *software for analysis of digital images*? This search engine lets us take that into account by specifying *image near analysis*. Now any pages with these words "near" each other will be detected. I want software (which might be called a program), and I don't want to pay for it. Thus, I would specify: **image and (processing or analysis) and (program or software) and (free or "no cost" or "public domain") and DNA and gel**. (Quotes around "public domain" cause it to be treated as one unit.)

While *search criteria* tells it what to look for, the *results ranking criteria* tells it how you want it to order the "hits" (sites determined to meet the search criteria). I used the same words (minus "and", "near" and "or") for ranking as used for searching. The program gives a "word count" for each of these words:

Word count: no cost: about 20000, gel:60897, public domain: about 80000, DNA:281594, processing:1060019, analysis: 1763973, free:2335512

Ignored: image: 4352192, software: 4894037, program: 5128465

Some words, like "software", "program" and "image" are used so frequently on Web pages that they don't really help much.

The program returns several pages of hits. Your mileage may vary, since the databases grow and change regularly, but when I did it, the first hit was a page with the enigmatic title "Chapter 2. Software Specifications". Following that link we find a description of a program called **NIH Image**, with a link to the file transfer protocol (FTP) site: <ftp://zippy.nimh.nih.gov/pub/nih-image/>

This server has the software, and the **README.TXT** file directs us to a Web page (<http://rsb.info.nih.gov/nih-image/>) for more information.

This search didn't locate the **NIH Image** homepage directly, but Web links got us close enough to find it. Several of the first hits link to **NIH Image**; if you hadn't suspected something called "Chapter 2" would have what you wanted, a few items further down the list you would see a page with the juicy title "Molecular Biology" (<http://alice.dcrf.nih.gov/molbio/index.html>). Following links from there gets you to **NIH Image** as well.

So, we are able to use the Internet both to find information about **NIH Image**, and to obtain a free copy! (U.S. tax dollars at work). **NIH Image** is only available as a Macintosh® version at present, but a Windows 95® image processing program called **UTHSCSA Image Tool** is available from the University of Texas Health Science Center at San Antonio (<http://ddsdx.uthscsa.edu/dig/itdesc.html>). I haven't tried it myself, but maybe somebody in Maine will find it useful.

Incidentally, the full text of *Walden* is available on the Internet at gopher://gopher.vt.edu:10010/02/149/5

BioBit: CDC Web Pages

The United States Centers for Disease Control has a collection of Web pages at: <http://www.cdc.gov/>

Pages are available for many diseases. For example, <http://www.cdc.gov/ncidod/diseases/vir/fvr/ebolainf.htm> covers the ever-popular Ebola Virus Hemorrhagic Fever.

This site also hosts the journal *Emerging Infectious Diseases*, published by the National Center for Infectious Diseases. Articles can be read or downloaded from: <http://www.cdc.gov/ncidod/EID/eid.htm>

The *Morbidity and Mortality Weekly Report* is available as portable document format (PDF) documents. PDF, a derivative of PostScript™, supplies documents as formatted pages. The free Adobe Acrobat™ PDF Reader, as well as instructions for setting up a variety of Web browsers to use Acrobat as a helper application, can be found at: <http://www.adobe.com/PDFs/online.pdf>

Versions are available for Macintosh, Windows™ and several flavors of UNIX™.